

## BIROn - Birkbeck Institutional Research Online

Levene, Mark and Fenner, Trevor and Bar-Ilan, J. (2019) Characterisation of the x-index and the rec-index. *Scientometrics* 120 (2), pp. 885-896. ISSN 0138-9130.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/27726/>

*Usage Guidelines:*

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>  
contact [lib-eprints@bbk.ac.uk](mailto:lib-eprints@bbk.ac.uk).

or alternatively

# Characterisation of the $\chi$ -index and the *rec*-index

Mark Levene mark@dcs.bbk.ac.uk<sup>1</sup>, Trevor Fenner trevor@dcs.bbk.ac.uk<sup>1</sup>, and  
Judith Bar-Ilan Judith.Bar-Ilan@biu.ac.il<sup>2</sup>

<sup>1</sup>Department of Computer Science and Information Systems,  
University of London, London WC1E 7HX, U.K.

<sup>2</sup>Department of Information Science, Bar-Ilan University, Ramat Gan, Israel

## Abstract

Axiomatic characterisation of a bibliometric index provides insight into the properties that the index satisfies and facilitates the comparison of different indices. A geometric generalisation of the *h*-index, called the  $\chi$ -index, has recently been proposed to address some of the problems with the *h*-index, in particular, the fact that it is not scale invariant, i.e., multiplying the number of citations of each publication by a positive constant may change the relative ranking of two researchers. While the square of the *h*-index is the area of the largest square under the citation curve of a researcher, the square of the  $\chi$ -index, which we call the *rec*-index (or *rectangle*-index), is the area of the largest rectangle under the citation curve. Our main contribution here is to provide a characterisation of the *rec*-index via three properties: *monotonicity*, *uniform citation* and *uniform equivalence*. Monotonicity is a natural property that we would expect any bibliometric index to satisfy, while the other two properties constrain the value of the *rec*-index to be the area of the largest rectangle under the citation curve. The *rec*-index also allows us to distinguish between *influential* researchers who have relatively few, but highly-cited, publications and *prolific* researchers who have many, but less-cited, publications.

*Keywords:* *h*-index,  $\chi$ -index, *rec*-index, bibliometric index, core publications, quantity versus quality, axiomatic characterisation

## 1 Introduction

Axiomatic characterisation of a bibliometric index [Mar09, BM14] provides insight into the properties that the index satisfies and facilitates the comparison of different indices. (Axiomatic characterisation is used in a number of areas for the same purpose.)

Since Garfield's foundational work in bibliometrics [Gar79], a plethora of bibliometric indices have been suggested [RB15, TB16], from a simple count of the total number of citations to the more sophisticated *h*-index [Hir05]. These have often been motivated by the ongoing debate between quantity (as measured by the number of publications) and quality (as measured by the number of citations to those publications). A review of many of these [WSL14], where a comparison of 108 bibliometric indicators was presented, concluded that, in order to gauge the overall impact of a researcher, several indicators should be used. In particular, many variants of the *h*-index have been proposed and, in [BMHD11], it was shown that generally there is high correlation between the *h*-index and 37 of its variants. Moreover,

a critical view of the  $h$ -index and its variants, which was provided in [Sch18], argued that it is not a good indicator of recent impact and suggested instead a timed  $h$ -index over a sliding time window. Another criticism of the  $h$ -index and its variants is that it treats all citations in the same way, ignoring the relevance and impact of citations. This issue can be addressed by Markovian methods, such as PageRank [Lev10], giving rise to indices based on author or publication-level citation networks, such as the Eigenfactor score [WJD<sup>+</sup>13], which can model influence between researchers and publications. However, Markovian indicators have the disadvantage of being significantly more complex to compute than those based on citation counts, and there is no conclusive evidence that they actually outperform rankings based on citation counts [FŠŽB15].

We mention two other issues concerning bibliometric indices that have been recently addressed in a formal setting. The first is that of normalising citation counts across different fields. For example, in [BM16], the authors investigated fractional counting of citations, where the value of a citation is inversely proportional to the numbers papers being cited. They provide a characterisation of the ranking induced by such fractional counting. The second issue is aggregation of bibliometric indices when there are several conflicting indices and there is no compelling reason to chose one over another. For example, in [SAP18], the authors investigated ranking methods from social choice theory in order to provide an axiomatic analysis of aggregation of bibliometric indices.

In this paper we focus on a (two-dimensional) geometric generalisation of the  $h$ -index, called the  $\chi$ -index [FHLB18], which has recently been proposed to address some of the problems with the  $h$ -index, in particular the fact that it is not *scale invariant* [PR16].

The  $h$ -index is determined by the largest square that fits under the citation curve of a researcher when plotting the number of citations to individual publications in decreasing order. On the other hand, the  $\chi$ -index is determined by the largest area rectangle that fits under the citation curve. The *rec*-index (or *rectangle*-index) is defined to be the area of this rectangle and is the square of the  $\chi$ -index. The *rec*-index is thus a member of the class of geometric indices that approximate the area under the citation curve of a researcher. Such indices can address problems attached to, for example, the citation count, which takes into account all citations. The  $h$ -index penalises both highly-cited publications and publications with only a few citations. On the other hand, the  $\chi$ -index is more balanced than the  $h$ -index in this respect, as it allows us to cater for both *influential* researchers with a few very highly-cited publications and *prolific* researchers who may have many publications but relatively few citations per publication.

Our main contribution here is to provide a first characterisation of the *rec*-index via three properties: *monotonicity*, *uniform citation* and *uniform equivalence*. Monotonicity is a natural property that we would expect any bibliometric index to satisfy, while the other two properties relate to the rectangle under the citation curve that determines the index. Uniform citation specifies that when the shape of the citation curve is uniform, i.e. rectangular, then the value of the index is the total citation count of all the publications. Complementing this property, uniform equivalence specifies that when the shape of the citation curve is not uniform, i.e. not rectangular, the value of the index is equal to that of some uniform citation curve that can be obtained by omitting some number of citations.

We note that many different properties may be used to characterise a bibliometric index, and there is no general agreement on which are the most compelling. For example, there are a number of distinct characterisations of the  $h$ -index, such as those presented in [Woe08b] and

[Que11a]. However, if a number of properties are proved to characterise a given index, the acceptance of any other index would necessitate the violation of at least one of the properties used to characterise the given index.

The rest of the paper is organised as follows. In Section 2, we introduce the *rec*-index. In Section 3, we define and discuss properties of the *rec*-index and other indices. Then, in Section 4, we present an axiomatic characterisation of the *rec*-index. Finally, in Section 5, we give our concluding remarks.

## 2 The *rec*-index and related bibliometric indices

We assume that a researcher publishes  $n$  publications, where  $n \geq 0$ , which are represented by a *citation vector* of *positive* integers,  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$ , where  $x_i$  is the number of citations to publication  $i$ , sorted in descending order, i.e.  $x_i \geq x_j$  for  $1 \leq i < j \leq n$ . (We note that one could consider only a subset of a researcher’s publications in the citation vector, for example, by only allowing journal publications, or publications in high-impact venues.)

The *citation curve* is the curve arising from plotting the number of citations against the ranking of the publications as a histogram specified by the citation vector.

A *bibliometric index* is a function  $f$  that maps citation vectors to the set of non-negative real numbers. As in [Woe08b], we assume the *baseline condition* that, for the empty citation vector  $\mathbf{x} = \langle \rangle$ , we have  $f(\mathbf{x}) = 0$ .

In this paper, we concentrate on characterising the *rec*-index directly, following the approach adopted in [Woe08b, Woe08a, Que11a, Que11b], rather than characterising the bibliometric ranking induced by the index, as was done in [Mar09, BM14]. This stems from our particular interest in the properties of geometric indices. It is evident that any two indices, such as the  $\chi$ -index and the *rec*-index, that are monotonic transformations of each other are equivalent with respect to the induced rankings.

The *citation count* index for a citation vector  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$  is the L1 norm of  $\mathbf{x}$ , denoted by  $\|\mathbf{x}\|$  and defined by

$$\|\mathbf{x}\| = \sum_{i=1}^n x_i. \quad (1)$$

We say that a citation vector  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$  is *dominated* by a citation vector  $\mathbf{y} = \langle y_1, y_2, \dots, y_m \rangle$ , written as  $\mathbf{x} \sqsubseteq \mathbf{y}$ , if  $n \leq m$  and  $x_i \leq y_i$  for all  $i$ ,  $1 \leq i \leq n$ .

It is worth recalling some known bibliometric indices [WSL14, TB16]: the citation count, as defined in (1); the publication count  $n$ ; the maximum citation index  $x_1$ ; the Euclidean index  $E(\mathbf{x})$  [PR16], which is the Euclidean norm of  $\mathbf{x}$ , i.e.,  $E(\mathbf{x}) = \sqrt{\sum_{i=1}^n x_i^2}$ ; the  $h$ -index [Hir05]; and the  $g$ -index [Egg06], which is a variant of the  $h$ -index giving extra weight to highly-cited publications.

The  $h$ -index [Hir05], in particular, has gained popularity due to its relative simplicity, ease of calculation, and its ingenious method of combining the quality and quantity of a researcher’s outputs. It is defined as the maximum number  $h$  of the researcher’s publications such that each has at least  $h$  citations, i.e. for a citation vector,  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$  the  $h$ -index is the largest  $h$  for which  $x_h \geq h$ .

To motivate the  $\chi$ -index, consider the following three citation vectors, the citation curves of which are depicted in Figure 1: (i)  $\langle 100 \rangle$ , i.e. 1 publication with 100 citations, (ii)  $\langle 10, 10, \dots, 10 \rangle$ , i.e. 10 publications with 10 citations each, and (iii)  $\langle 1, 1, \dots, 1 \rangle$ , i.e. 100 publications with 1 citation each. (Note that the diagram in Figure 1 is not drawn to scale.)

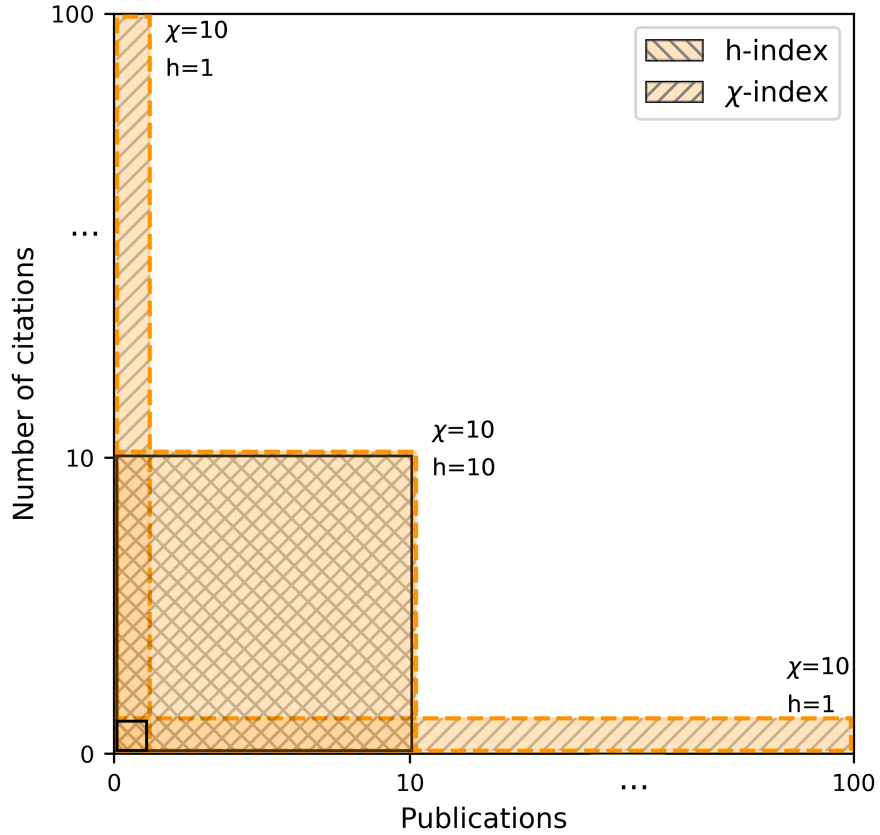


Figure 1: Example of the geometric interpretation of the  $h$  and  $\chi$  indices.

The  $\chi$ -index is defined as the square root of the maximum area rectangle that can fit under the citation curve, while the  $h$ -index is the square root of the maximum area square that can fit under the citation curve.

Formally, we first define the *rec*-index (or *rectangle*-index) of a researcher with citation vector  $\mathbf{x}$  by

$$rec(\mathbf{x}) = \max_i x_i. \quad (2)$$

The  $\chi$ -index [FHLB18] is then defined by  $\chi(\mathbf{x}) = \sqrt{rec(\mathbf{x})}$ .

Returning to our example shown in Figure 1, we see that all three researchers have a  $\chi$ -index of 10, while researcher (ii) has an  $h$ -index of 10, but researchers (i) and (iii) both have an  $h$ -index of only 1. The  $h$ -index may be seen as balancing quality, on the one hand, by

favouring publications with a higher number of citations and quantity, on the other hand, by taking into account all publications with a sufficient number of citations. However, such an approach disadvantages a researcher, such as (i), with a few very highly-cited publications, who may have carried out some influential seminal research, and it also disadvantages a prolific researcher, such as (iii), who may have many publications but fewer citations per publication. Now, if we let  $k$  denote a value of  $i$  that maximises  $ix_i$  in (2), the *rec*-index can distinguish between more *influential* researchers for which  $x_k > k$ , such as (i), and more *prolific* researchers for which  $k > x_k$ , such as (iii). In this sense the  $\chi$ -index avoids the debate of number of citations versus number of publications by awarding all three researchers the same  $\chi$ -index of 10.

The *rec*-index is a member of the class of (two-dimensional) geometric indices, as is the square of the *h*-index, and also the half the square of the *w*-index [Woe08b], which is the area of the maximal isosceles right-angled triangle under the citation curve; formally, the *w*-index is the largest integer  $w$  such that the citation vector  $\mathbf{x}$  contains  $w$  distinct publications with at least  $1, 2, \dots, w$  citations, respectively. In a more general setting, the dimension of an index can be formally defined [Pra17], and is related to the requirement of *dimensional homogeneity* from physics, that we may only compare quantities that have the same units.

Geometric indices are actually quite natural, as their goal is to consider the area under the citation curve in order to encapsulate the essential citations for a set of *core* publications that in some sense represent the output of a researcher. We note that the *rec*-index, the square of the *h*-index, the publication count, and the maximum citation index all include the same number of citations for each core publication.

The citation count, which includes all publications with at least one citation in the core, is of course a reasonable bibliometric index. However, it is often argued that the citation count is problematic; in particular, it may be inflated by a small number of publication having a large number of citations or it may be sensitive to a long tail of publications each having only few citations. The axiomatic characterisation we describe here can be viewed as contributing to this debate by discussing several characteristics of geometric indices.

### 3 Some properties of bibliometric indices

In this section we define a variety of properties of bibliometric indices, almost all of which are satisfied by the *rec*-index, and then, in Section 4, we show that *monotonicity* combined with *uniform citation* and *uniform equivalence* characterise the *rec*-index.

We assume throughout this section that  $f(\cdot)$  is the index under consideration. The first property, monotonicity, is a natural requirement for any bibliometric index, stating that adding citations to the citation vector should not decrease the value of the index:

**Monotonicity (M):** If citation vector  $\mathbf{x}$  is dominated by citation vector  $\mathbf{y}$ , i.e.  $\mathbf{x} \sqsubseteq \mathbf{y}$ , then  $f(\mathbf{x}) \leq f(\mathbf{y})$ .

It is easy to verify that all the indices we consider are monotonic. We note that the citation count satisfies the stronger property of *strict monotonicity* (**SM**), in which  $f(\mathbf{x}) < f(\mathbf{y})$  when  $\mathbf{x} \neq \mathbf{y}$ . It also satisfies several other desirable properties, such as *rank independence* (adding a new publication with a given number of citations to two citation vectors does not change their relative ranking) and *rank scale invariance* (multiplying the number of citations of each

publication by a positive constant does not change the relative ranking of two citation vectors) [PR16]. Neither the  $\chi$ -index nor the  $h$ -index are rank independent. However, the  $\chi$ -index is rank scale invariant, but the  $h$ -index is not. Moreover, the *rec*-index satisfies the following stronger property of (linear) scale invariance:

**Scale invariance (SI):**  $f(C\mathbf{x}) = Cf(\mathbf{x})$ , for any positive constant  $C$ .

It is easy to see that *scale invariance* implies rank scale invariance.

A natural form of symmetry can be attained via the *conjugate partition* of a citation vector  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$ , which is the *publication vector*  $\mathbf{p} = \langle p_1, p_2, \dots, p_m \rangle$ , where  $m = x_1$  and  $p_i$  is the number of publications with at least  $i$  citations [Woe08b]. Geometrically, the publication vector is obtained by reflecting the geometric representation of the citation vector along the main diagonal. This is shown in Figure 2 for a citation vector  $\langle 6, 4, 3, 1 \rangle$ , on the left, and its conjugate partition publication vector  $\langle 4, 3, 3, 2, 1, 1 \rangle$ , on the right. This motivates the following property.

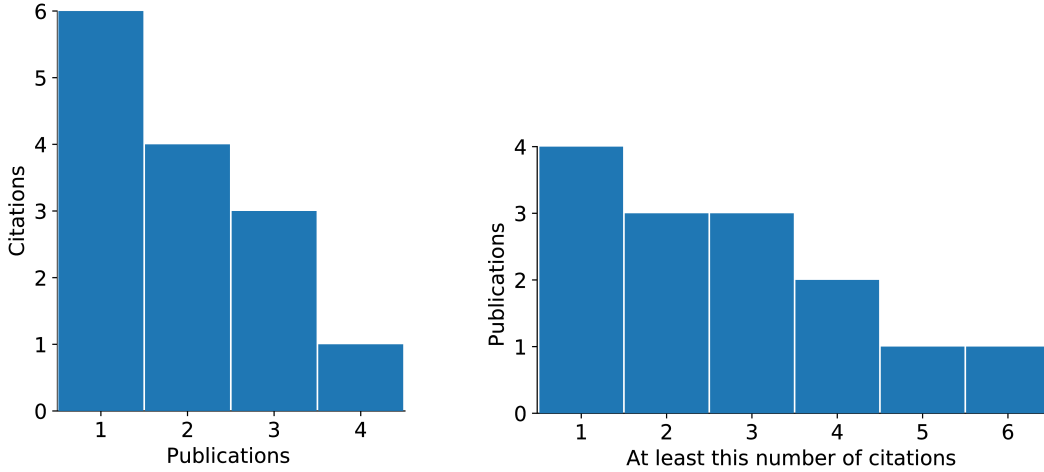


Figure 2: The citation vector is shown on the left and its conjugate partition, the corresponding publication vector, is shown on the right.

**Self-conjugacy (SC):** Let  $\mathbf{p}$  be the conjugate partition of the citation vector  $\mathbf{x}$ , then  $f(\mathbf{p}) = f(\mathbf{x})$ .

Clearly the *rec* and  $h$  indices, as well as the citation count, are all self-conjugate. On the other hand, we note that the publication count and the maximum citation index are conjugates of each other. Self-conjugacy implies a balanced approach between influence (quality) and prolificity (quantity).

Some indices tend to emphasise influence, for example, the maximum citation index, Euclidean index and  $g$ -index, whereas others, such as the publication count, emphasise prolificity. Should we wish to emphasise influence rather than prolificity, we may define a version of the *rec*-index, the *rec<sub>i</sub>*-index, in which we restrict the maximum in (2) to be over those  $i$  for which  $i \leq x_i$ . Conversely, should we wish to emphasise prolificity, we may instead define

the conjugate index, the  $rec_P$ -index, by using the corresponding publication vector  $\mathbf{p}$  and restricting  $i$  so that  $i \leq p_i$ .

Many people, probably the majority, tend to favour indices that emphasise influence. It is therefore worth noting our findings in [FHLB18], where the citations of a large number of researchers, from a Google Scholar data set made available by Radicchi and Castellano [RC13], were analysed and their  $rec$ -indices calculated. Table 11 in [FHLB18] shows that 93% of the researchers for which the  $\chi$ -index was significantly larger than the  $h$ -index were more influential than prolific, i.e. for these researchers  $x_k > k$ . This indicates that, in general, the  $rec$ -index satisfies the tendency to favour influence.

Following the conclusion in [WSL14] that several bibliometric indicators should be used to gauge the overall impact of a researcher, we suggest that both quality and quantity may be assessed by using the pair of indices  $(rec_I, rec_P)$ .

We now concentrate on the  $rec$ -index. A typical citation curve, corresponding to the citation vector  $\langle 6, 4, 3, 1 \rangle$ , is shown in Figure 3, with circles indicating where a new citation can be added. We observe that each addition of a new citation completes a rectangle. If the new citation is added to publication  $k$ , then the newly formed rectangle has width  $k$  and height  $x_k + 1$ . For example, if we add a citation to the third publication, producing the citation vector  $\langle 6, 4, 4, 1 \rangle$ , then the  $rec$ -index will increase from 9 to 12, but if we add it to the fourth publication, producing the citation vector  $\langle 6, 4, 3, 2 \rangle$ , then the  $rec$ -index will not increase.

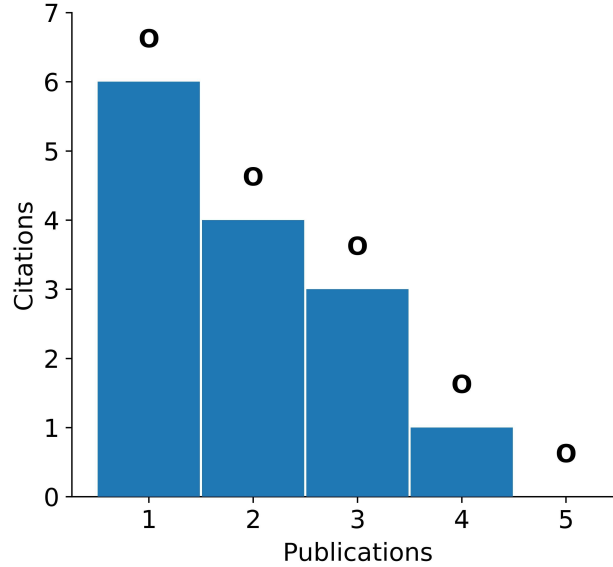


Figure 3: A citation vector with circles indicating where a new citation may be added.

For any citation vector  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$ , we write  $\mathbf{x}^{[k]}$  for the citation vector obtained from  $\mathbf{x}$  by adding a single citation to publication  $k$ , where  $1 \leq k \leq n + 1$ , thereby increasing its citation count from  $x_k$  to  $x_k + 1$ . If  $k = n + 1$ , we assume that  $x_k = 0$ . We note that  $k$  must be the smallest index  $j$  for which  $x_j = x_k$ .

By the definition of the  $rec$ -index, it is straightforward to see that  $rec(\mathbf{x}^{[k]}) \geq k(x_k + 1)$ ,



and thus

$$\text{rec}(\mathbf{x}^{[k]}) = \max(\text{rec}(\mathbf{x}), k(x_k + 1)). \quad (3)$$

The following property encapsulates this observation.

**Rectangle completion (RC):** For any citation vector  $\mathbf{x}$ ,

$$f(\mathbf{x}^{[k]}) = \max(f(\mathbf{x}), k(x_k + 1)).$$

Noting that *rectangle completion* implies *monotonicity*, it is then straightforward to verify that a bibliometric index  $f$  satisfies *rectangle completion* (together with the baseline condition) if and only if it is the *rec*-index. Obviously, as it essentially encapsulates the definition of the *rec*-index, *rectangle completion* is not particularly useful as a characterisation of the *rec*-index; however, it provides an alternative and constructive definition of the index.

Consider the situation when  $\text{rec}(\mathbf{x}) = kx_k$ . We note that (i) if  $x_k = k$ , the *rec*-index is equal to the square of the *h*-index, (ii) if  $x_k > k$ , the researcher tends towards being more influential, and (iii) if  $x_k < k$ , the researcher tends towards being more prolific. Thus the shape of the maximum area rectangle will indicate an interpretation of the index value. We also note that, when the histogram of the citation curve is a rectangle, the distribution of citations is *uniform*. In this case the *core* includes all publications.

More formally, we say that a citation vector  $\mathbf{u} = \langle u_1, u_2, \dots, u_n \rangle$  is *uniform* if  $u_1 = u_2 = \dots = u_n$ . It follows that  $\text{rec}(\mathbf{x}) = \|\mathbf{x}\|$  if and only if  $\mathbf{x}$  is uniform. This observation suggests the following weaker form of this property.

**Uniform citation (UC):** If the citation vector  $\mathbf{x}$  is uniform then  $f(\mathbf{x}) = \|\mathbf{x}\|$ .

This property makes the reasonable assertion that, if all publications have the same number of citations, they should all be in the core and all citations to them included in the index.

As stated above *rectangle completion* is too contrived in the sense that it mimics the definition of the *rec*-index. So we now explore a way to replace it with *uniform citation*, which gives a lower bound on the index, together with another property that gives an upper bound on the index. This additional property is:

**Uniform equivalence (UE):** For any citation vector  $\mathbf{x}$ , there exists a uniform citation vector  $\mathbf{u}$  dominated by  $\mathbf{x}$ , i.e.  $\mathbf{u} \sqsubseteq \mathbf{x}$ , for which  $f(\mathbf{x}) = f(\mathbf{u})$ .

This property asserts that the same number of citations should be included for each publication in the core.

The following proposition, which defines a property that is similar to Axiom D in [Woe08b], can easily be shown to follow from *monotonicity*, *uniform citation* and *uniform equivalence*.

**Proposition 3.1.** *If  $f$  satisfies properties M, UC and UE then it also satisfies the following property:*

**Citation increase (CI):** *If we add a single citation to each publication in  $\mathbf{x}$ , resulting in the citation vector  $\mathbf{y}$ , then  $f(\mathbf{y}) > f(\mathbf{x})$ .*

However, *citation increase* together with *monotonicity* and *uniform citation* do not imply *uniform equivalence*, as is the case for the bibliometric index,

$$f(\mathbf{x}) = \frac{\text{rec}(\mathbf{x}) + \|\mathbf{x}\|}{2}, \quad (4)$$

which satisfies *citation increase* but not *uniform equivalence*.

## 4 Axiomatic characterisation of the *rec*-index

The rankings induced by several bibliometric indices, including the *h*-index and *g*-index, were characterised in [Mar09, BM14], whereas, in [Woe08b, Woe08a] and [Que11a, Que11b], the authors concentrated on characterising the *h*-index and *g*-index directly. These characterisations, as well as that presented here, are based on properties that address three types of issues. First, the inclusion of fundamental properties like baseline and monotonicity should be considered. The second issue is concerned with the conditions under which the value of the index increases (see, for example, *citation increase*). The third issue considers what changes to the citation vector leave the index unchanged (for example, the *h*-index satisfies the property of *independence of irrelevant citations* [Que11b], cf. [Kon14], which essentially states that adding a single citation to a core publication does not increase the index). Another category of properties that are important for characterising bibliometric indices are invariants, such as *scale invariance* and *self-conjugacy*, which give transformations that change the index in a predictable manner or do not change the value at all.

The main result in this section is Theorem 4.1, which provides an axiomatic characterisation of the *rec*-index. We also show in Proposition 4.5 that the  $\chi$ -index satisfies the desirable property that, subsequent to the addition of a single citation to the citation vector, the  $\chi$ -index cannot increase by more than one.

**Theorem 4.1.** *A bibliometric index  $f$  satisfies the three properties of monotonicity, uniform citation and uniform equivalence if and only if it is the *rec*-index.*

*Proof.* It is clear from Section 3 that the *rec*-index satisfies these properties, so it remains to prove that they are sufficient.

Let  $\mathbf{x}$  be a citation vector and let  $\mathbf{u}$  be a uniform citation vector such that  $\|\mathbf{u}\|$  is maximal for all  $\mathbf{u} \sqsubseteq \mathbf{x}$ . Clearly,  $\text{rec}(\mathbf{x}) = \|\mathbf{u}\|$ . So, by *monotonicity* and *uniform citation*, we have  $f(\mathbf{x}) \geq f(\mathbf{u}) = \|\mathbf{u}\|$ . By *uniform equivalence*, we may let  $\mathbf{v}$  be a uniform citation vector such that  $\mathbf{v} \sqsubseteq \mathbf{x}$  and  $f(\mathbf{x}) = f(\mathbf{v})$ , and therefore  $f(\mathbf{x}) = \|\mathbf{v}\|$  by *uniform citation*. Since  $\|\mathbf{u}\| \geq \|\mathbf{v}\|$  by the definition of  $\mathbf{u}$ , it follows that  $f(\mathbf{x}) = \|\mathbf{u}\| = \text{rec}(\mathbf{x})$ .  $\square$

The following two corollaries show that we may replace two of the properties in Theorem 4.1 by two simpler properties together with the intuitive property of *scale invariance*.

**Corollary 4.2.** *The result of Theorem 4.1 holds if we replace monotonicity by the following more restrictive property.*

**Uniform monotonicity (UM):** *If  $\mathbf{x}$  is uniform and  $\mathbf{x} \sqsubseteq \mathbf{y}$ , then  $f(\mathbf{x}) \leq f(\mathbf{y})$ .*

**Corollary 4.3.** *The result of Theorem 4.1 holds if we replace uniform citation by the following more restrictive property together with scale invariance.*

**Uniform single citation (USC):** *If  $x_1 = x_2 = \dots = x_n = 1$  for the citation vector  $\mathbf{x}$ , then  $f(\mathbf{x}) = \|\mathbf{x}\| = n$ .*

*Proof.* Clearly **USC** and **SI** imply **UC**. □

It is not difficult to demonstrate that the three properties of Theorem 4.1 characterising the *rec*-index are independent, i.e. omitting any one of them would render the theorem false.

**Proposition 4.4.** *The three properties of Theorem 4.1 characterising the *rec*-index are independent.*

*Proof.* The following examples justify this claim.

- a) The index defined in (4) satisfies **M** and **UC** but not **UE**.
- b) The square of the *h*-index, the publication count  $n$ , the maximum citation index  $x_1$ ,  $\max(n, x_1)$  and  $\min(n, x_1)$  all satisfy **M** and **UE** but not **UC**.
- c) The product  $nx_n$  of the number of publications and the minimum number of citations satisfies **UC** and **UE** but not **M**, nor **UM**.

□

To summarise the properties that characterise the *rec*-index: **UM** or **M** implies that adding new citations will not decrease the value of the index, while **UC** and **UE** provide, respectively, lower and upper bounds on its value. We recall that, since the *rec*-index is not strictly monotonic, some citations may not contribute to the value of the index; however, all core publications contribute an equal number of citations.

We now present a construction that gives rise to a property equivalent to **UE**. While indices, such as the citation count (1), that include the full set of citations are strictly monotonic, indices, such as the *rec*-index, that include just a core set of publications typically increase only after a batch of citations has been added. The construction we now present captures this aspect of the *rec*-index.

Consider a sequence of citation vectors

$$\mathbf{S} = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_s, \tag{5}$$

where  $\mathbf{x}_1 = \langle \rangle$  and  $\mathbf{x}_s = \mathbf{x}$ . When  $\mathbf{x}_i \subseteq \mathbf{x}_{i+1}$  and  $\|\mathbf{x}_{i+1}\| - \|\mathbf{x}_i\| = 1$ , for  $1 \leq i \leq s-1$ , we say that  $\mathbf{S}$  is a *constructive* sequence for  $\mathbf{x}$ .

We are interested in constructive sequences satisfying the property that, for each  $i$ , if  $f(\mathbf{x}_i) < f(\mathbf{x}_{i+1})$  then  $\mathbf{x}_{i+1}$  is a uniform citation vector; we call such a constructive sequence *f-incremental*. This suggests the following property of a citation index.

**Uniform increment (UI):** For any citation vector  $\mathbf{x}$ , there exists an *f-incremental* constructive sequence for  $\mathbf{x}$ .

It is not difficult to prove that **UI** implies **UE**. Moreover, the *rec*-index satisfies **UI**, and we now present one method for constructing a *rec-incremental* sequence for a citation vector  $\mathbf{x}$ .

- (i) Start from  $\langle \rangle$  and construct a sequence of uniform citation vectors as follows.
- (ii) From the uniform citation vector  $\mathbf{u}$ , add citations one-by-one to obtain a new uniform citation vector, either by adding a new column (i.e. a new publication) or by adding a new row (cf. property **CI**) to the rectangle corresponding to  $\mathbf{u}$ .
- (iii) Repeat step (ii) until we obtain a citation vector  $\mathbf{u}$  for which  $\text{rec}(\mathbf{u}) = \text{rec}(\mathbf{x})$ .
- (iv) Add the remaining citations one-by-one in any order until we obtain  $\mathbf{x}$ .

It is straightforward to show that it is always possible, in step (ii) above, to choose between adding a column or a row in such a way that the sequence is *rec-incremental*.

It may be argued that it is natural that a one-dimensional bibliometric index should not increase by more than one when a single citation is added to the citation vector. We now prove that this holds for the  $\chi$ -index.

**Proposition 4.5.** *Let  $\mathbf{x}$  be a citation vector and let  $\mathbf{x}^{[k]}$  be a citation vector obtained by adding a single citation to  $\mathbf{x}$ . Then  $\chi(\mathbf{x}^{[k]}) \leq \chi(\mathbf{x}) + 1$ .*

*Proof.* We recall that  $\mathbf{x}^{[k]}$  is obtained from  $\mathbf{x}$  by adding a single citation to publication  $k$ , thereby increasing its citation count to  $x_k + 1$ . If  $\text{rec}(\mathbf{x}^{[k]}) = \text{rec}(\mathbf{x})$  the result holds trivially. So we may assume that  $\text{rec}(\mathbf{x}^{[k]}) \neq \text{rec}(\mathbf{x})$ , in which case  $\text{rec}(\mathbf{x}^{[k]}) = k(x_k + 1)$  by (3).

Now, since  $\mathbf{x}^{[k]}$  is a citation vector, we must have

$$\text{rec}(\mathbf{x}) \geq \max(kx_k, (k-1)(x_k+1)).$$

Therefore,

$$\text{rec}(\mathbf{x}) \geq \text{rec}(\mathbf{x}^{[k]}) - \min(k, x_k + 1) \geq \text{rec}(\mathbf{x}^{[k]}) - \sqrt{k(x_k + 1)},$$

and thus

$$\text{rec}(\mathbf{x}^{[k]}) - \text{rec}(\mathbf{x}) \leq \chi(\mathbf{x}^{[k]}).$$

It follows that

$$\chi(\mathbf{x}^{[k]}) - \chi(\mathbf{x}) \leq \frac{\chi(\mathbf{x}^{[k]})}{\chi(\mathbf{x}^{[k]}) + \chi(\mathbf{x})} \leq 1.$$

□

Finally, we note that if it is required that the  $\chi$ -index be an integer, then the ceiling function, which maps  $\chi(\mathbf{x})$  to the least integer greater than or equal to  $\chi(\mathbf{x})$ , can be employed.

## 5 Concluding remarks

Geometric indices, such as the *rec*-index, capture a core set of the publications that represent a researcher’s total output. The  $\chi$ -index (which is equal to the square root of the *rec*-index) can be viewed as a generalisation of the *h*-index, and has the advantage that it allows us to distinguish between more influential and more prolific researchers, depending on whether the height of the largest area rectangle under the citation curve is greater than or less than its width, respectively.

We presented several properties that are satisfied by the *rec*-index and proved, in Theorem 4.1, that the three properties of *monotonicity*, *uniform citation* and *uniform equivalence* characterise the *rec*-index. While monotonicity is a very natural property for any bibliometric index, uniform citation and uniform equivalence are natural when it is required to include the same number of citations for each publication in the core.

Geometric indices, such as the *rec*-index, give better insight into the relationship between influence (quality) and prolificity (quantity) than indices, such as the *h*-index, that are more constrained in this respect.

## Acknowledgements

The authors would like to thank the reviewers for their constructive comments, which have helped us to improve the paper.

## References

- [BM14] D. Bouyssou and T. Marchant. An axiomatic approach to bibliometric rankings and indices. *Journal of Informetrics*, 8:449–477, 2014.
- [BM16] D. Bouyssou and T. Marchant. Ranking authors using fractional counting of citations: An axiomatic approach. *Journal of Informetrics*, 10:183–199, 2016.
- [BMHD11] L. Bornmann, R. Mutz, S.E. Hug, and H.-D. Daniel. A multilevel meta-analysis of studies reporting correlations between the *h* index and 37 different *h* index variants. *Journal of Informetrics*, 5:346–359, 2011.
- [Egg06] L. Egghe. Theory and practise of the *g*-index. *Scientometrics*, 69:131–152, 2006.
- [FHLB18] T. Fenner, M. Harris, M. Levene, and J. Bar-Ilan. A novel bibliometric index with a simple geometric interpretation. *PLoS ONE*, 13(7):e0200098, 2018.
- [FŠŽB15] D. Fiala, L. Šubelj, S. Žitnik, and M. Bajec. Do PageRank-based author rankings outperform simple citation counts? *Journal of Informetrics*, 9:334–348, 2015.
- [Gar79] E. Garfield. Is citation analysis a legitimate evaluation tool? *Scientometrics*, 1:359–375, 1979.
- [Hir05] J.E. Hirsch. An index to quantify an individual’s scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 98:16569–16572, November 2005.

- [Kon14] T. Kongo. An alternative axiomatization of the Hirsch index. *Journal of Informetrics*, 8:252–258, 2014.
- [Lev10] M. Levene. *An Introduction to Search Engines and Web Navigation*. John Wiley & Sons, Hoboken, New Jersey, second edition, 2010.
- [Mar09] T. Marchant. An axiomatic characterization of the ranking based on the h-index and some other bibliometric rankings of authors. *Scientometrics*, 80:325–342, 2009.
- [PR16] M. Perry and P.J. Reny. How to count citations if you must. *The American Economic Review*, 106:2722–2741, 2016.
- [Pra17] G. Prathap. Citation indices and dimensional homogeneity. *Current Science*, 113:853–855, 2017.
- [Que11a] A. Quesada. Axiomatics for the Hirsch index and the Egghe index. *Journal of Informetrics*, 5:476–480, 2011.
- [Que11b] A. Quesada. Further characterizations of the Hirsch index. *Scientometrics*, 87:107–114, 2011.
- [RB15] R.C. Roemer and R. Borchardt. *Meaningful Metrics: A 21st-Century Librarians Guide to Bibliometrics, Altmetrics, and Research Impact*. Association of College and Research Libraries, Chicago, Il., 2015.
- [RC13] F. Radicchi and C. Castellano. Analysis of bibliometric indicators for individual scholars in a large data set. *Scientometrics*, 97:627–637, 2013.
- [SAP18] A. Subochev, F. Aleskerov, and V. Pislyakov. Ranking journals using social choice theory methods: A novel approach in bibliometrics. *Journal of Informetrics*, 12:416–429, 2018.
- [Sch18] M. Schreiber. A skeptical view on the Hirsch index and its predictive power. *Physica Scripta*, 93(10):102501, 2018.
- [TB16] R. Todeschini and A. Baccini. *Handbook of Bibliometric Indicators: Quantitative Tools for Studying and Evaluating Research*. Wiley-VCH, Weinheim, Germany, 2016.
- [WJD<sup>+</sup>13] J.D. West, M.C. Jensen, R.J. Dandrea, G.J. Gordon, and C.T. Bergstrom. Author-level Eigenfactor metrics: Evaluating the influence of authors, institutions, and countries within the social science research network community. *Journal of the American Society for Information Science and Technology*, 64:787–801, 2013.
- [Woe08a] G.J. Woeginger. An axiomatic analysis of Egghe’s  $g$ -index. *Journal of Informetrics*, 2:364–368, 2008.
- [Woe08b] G.J. Woeginger. A symmetry axiom for scientific impact indices. *Journal of Informetrics*, 2:298–303, 2008.

- [WSL14] L. Wildgaard, J.W. Schneider, and B. Larsen. A review of the characteristics of 108 author-level bibliometric indicators. *Scientometrics*, 101:125–158, 2014.